



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

## A high-resolution single nucleotide polymorphism genetic map of the mouse genome

**Citation for published version:**

Shifman, S, Bell, JT, Copley, RR, Taylor, MS, Williams, RW, Mott, R & Flint, J 2006, 'A high-resolution single nucleotide polymorphism genetic map of the mouse genome', *PLoS Biology*, vol. 4, no. 12, pp. e395. <https://doi.org/10.1371/journal.pbio.0040395>

**Digital Object Identifier (DOI):**

[10.1371/journal.pbio.0040395](https://doi.org/10.1371/journal.pbio.0040395)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Publisher's PDF, also known as Version of record

**Published In:**

PLoS Biology

**Publisher Rights Statement:**

This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



# A High-Resolution Single Nucleotide Polymorphism Genetic Map of the Mouse Genome

Sagiv Shifman<sup>1</sup>, Jordana Tzenova Bell<sup>1</sup>, Richard R. Copley<sup>1</sup>, Martin S. Taylor<sup>1</sup>, Robert W. Williams<sup>2</sup>, Richard Mott<sup>1</sup>, Jonathan Flint<sup>1\*</sup>

**1** Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom, **2** Department of Anatomy and Neurobiology, University of Tennessee Health Science Center, Memphis, Tennessee, United States of America

**High-resolution genetic maps are required for mapping complex traits and for the study of recombination. We report the highest density genetic map yet created for any organism, except humans. Using more than 10,000 single nucleotide polymorphisms evenly spaced across the mouse genome, we have constructed genetic maps for both outbred and inbred mice, and separately for males and females. Recombination rates are highly correlated in outbred and inbred mice, but show relatively low correlation between males and females. Differences between male and female recombination maps and the sequence features associated with recombination are strikingly similar to those observed in humans. Genetic maps are available from [http://gscan.well.ox.ac.uk/#genetic\\_map](http://gscan.well.ox.ac.uk/#genetic_map) and as supporting information to this publication.**

Citation: Shifman S, Bell JT, Copley RR, Taylor MS, Williams RW, et al. (2006) A high-resolution single nucleotide polymorphism genetic map of the mouse genome. PLoS Biol 4(12): e395. DOI: 10.1371/journal.pbio.0040395

## Introduction

The current genetic map of the mouse is based on data from crosses between inbred strains constructed using simple sequence length polymorphism markers in the late 1990s [1,2]. The most recent, and still the most comprehensive map, was created by backcrossing SPRET/EI (*Mus spretus*) to C57BL/6 and mapping the position of 3,368 markers on 982 progeny. The map consists of 2,302 genetically separated bins, with an average distance between bins of 0.61 centimorgans (cM) [3].

The completion of the mouse genome [4] together with the discovery of thousands of single nucleotide polymorphisms (SNPs) [5–7] has meant that high-resolution maps can be simply obtained by converting physical to genetic distances, since on average 1 cM corresponds to 1.6 megabases (Mb). However, although any number of SNPs can be combined to form a map using this approach [8], recombination rates are not constant across the genome [9,10], so at high resolution, the inferred genetic distances will be incorrect.

There are a number of applications for which accurate distances are needed. First, at high resolution, recombination-based genetic mapping of phenotypic variation will incur errors if incorrect genetic map distances are used. Recombination-based mapping is an efficient method to tackle a major challenge facing complex trait genetics: mapping quantitative trait loci (QTLs) into sufficiently small intervals to make gene identification possible. By exploiting historical recombinants that have accumulated over many generations, either in a genetically heterogeneous stock (HS) of mice descended from eight inbred progenitor strains [11,12], or in an advanced intercross [13], QTLs can be mapped into intervals of a few megabases. The accuracy of these methods depends on the accuracy of the genetic map. Additionally, as in human linkage mapping, a substantial difference in the number of informative male and female meioses will decrease power [14,15].

Second, accurate genetic maps are important for investigating recombination hotspots and the causes of recombination rate variation across the mammalian genome. For human populations, it has been possible to estimate fine-scale recombination rates from linkage-disequilibrium (LD) data, using a coalescent-based method and certain assumptions about population history. These data support the hypothesis of a two-stage process for recombination in which recombination rates are constrained over large scales, but rapidly change over a fine scale [10,16]. Consistent with this view, the location of hotspots rarely occurs in the same positions in humans and chimpanzees [17]. However, the features that induce and maintain recombination hotspots in mammals remain largely obscure. A high-resolution recombination-rate map in the mouse would make it possible to investigate these phenomena in a genetically tractable model organism.

We set out to construct a very high-resolution genetic map using more than 10,000 SNPs. One map was constructed using eight recombinant inbred (RI) lines (AXB, BXA, CXB, BXD, BXH, AKXD, LXS, and SWXJ). The second was constructed

**Academic Editor:** Richard Gibbs, Baylor College of Medicine, United States of America

**Received:** June 23, 2006; **Accepted:** September 18, 2006; **Published:** November 14, 2006

**DOI:** 10.1371/journal.pbio.0040395

**Copyright:** © 2006 Shifman et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Abbreviations:** cM, centimorgan; HS, heterogeneous stock; kb, kilobase; LD, linkage disequilibrium; LINE, long interspersed nuclear element; LOD, log of odds; Mb, megabase; NCBI, National Center for Biotechnology Information; QTL, quantitative trait locus; RI, recombinant inbred; RR, relative risk; SD, standard deviation; SNP, single nucleotide polymorphism

\* To whom correspondence should be addressed. E-mail: [jf@well.ox.ac.uk](mailto:jf@well.ox.ac.uk)

**Table 1.** Description of the SNP Genetic Map in the HS Pedigrees

Chromosome	SNPs	Map Positions <sup>a</sup>	Genetic Length (cM)			Map Resolution <sup>b</sup>	Physical Length <sup>c</sup> (Mb)	cM/Mb
			Sex Averaged	Female	Male			
1	873	388	118.1	133.5	103.5	0.3	194.9	0.61
2	762	331	108.4	121.9	97.0	0.33	182.6	0.59
3	698	318	89.6	104.4	73.9	0.28	159.9	0.56
4	678	344	101.7	114.7	84.9	0.3	155.2	0.66
5	597	286	106.8	111.1	98.4	0.37	153.1	0.70
6	611	294	90.0	101.4	79.7	0.31	149.7	0.60
7	512	262	89.7	95.4	87.6	0.34	141.8	0.63
8	520	213	80.5	96.7	65.6	0.38	127.9	0.63
9	441	231	86.0	96.8	75.0	0.37	123.8	0.70
10	359	181	83.4	83.6	83.4	0.46	130.1	0.64
11	614	293	97.1	101.3	92.6	0.33	122.1	0.80
12	465	239	69.3	75.1	66.0	0.29	117.8	0.59
13	454	206	69.5	82.7	57.6	0.34	116.7	0.60
14	489	182	59.7	71.9	50.6	0.33	119.2	0.50
15	454	213	65.1	72.3	60.8	0.31	103.7	0.63
16	420	197	63.2	77.0	49.2	0.32	98.5	0.64
17	398	166	63.0	66.5	59.8	0.38	93.3	0.68
18	338	171	64.2	85.3	48.2	0.38	90.9	0.71
19	221	121	54.1	55.9	51.9	0.45	61.2	0.88
X	298	89	70.5	70.5	-	0.79	164.9	0.43
Total	10,202	4,725	1,629.8	1,817.4	1,385.7	0.37	2,607.3	0.63

<sup>a</sup>The number of unique positions of the sex-averaged map.<sup>b</sup>The length of the sex-averaged map divided by the number of unique map positions.<sup>c</sup>Physical length is the distance between the first and last SNP.

DOI: 10.1371/journal.pbio.0040395.t001

using recombinants that have accumulated over four generations in a genetically heterogeneous stock of outbred mice [18]. The HS are descended from eight inbred progenitors (AKR/J, A/J, BALB/c, CBA/J, C3H/HeJ, C57BL/6J, DBA/2J, and LP/J), and although their genomes are fine-grained mosaics of these founders, for the purpose of constructing the genetic map, we ignored their ancestry and focused only on the last four generations of the cross for which full pedigree information was available.

## Results

### Genotype Quality and Accuracy

A total of 13,367 SNPs were genotyped using the Illumina BeadArray platform (Illumina, San Diego, California, United States) with call rates of 99.86%. The average interval size between the selected SNPs based on National Center for Biotechnology Information (NCBI) build 34 is 167.0 kilobases (kb) (standard deviation [SD] = 141.5); 99.0% of the intervals are below 500 kb and 81.2% are below 250 kb. Genotype accuracy, based on the inheritance of alleles in 1,097 trios and 25 duplicate samples, is 99.99%. From 5,089 SNPs typed on 42 inbred strains by us and by Merck [19], there was 0.041% disagreement. Assuming an equal error rate in the two studies, the estimated accuracy of our data is 99.98%.

### Genetic Maps

Maps were created separately for HS mice and RI strains. Genotypes were obtained for 11,247 informative SNP markers in 2,293 HS animals. There are 84 HS pedigrees, eight of which are large, complex multi-generational families with inbreeding loops, encompassing the majority (61.1%) of individuals. The remaining 72 families are nuclear, with an

average sibship size of 9.6 (range 2–34). The pedigrees consisted of 4,048 potentially informative meioses, of which the average number of informative meioses per marker was 1,296.7 (range: 9–2,755 per marker), and the average number of phase-known meioses was 198.1 (range: 7–964 per marker).

We estimated genetic distances using the software program CRIMAP [20]. Initially we took the physical order of marker from NCBI build 34 of the mouse genome and required each marker to exhibit linkage to at least one other marker on its chromosome. Because of the complexity of the HS pedigrees and the number of markers, it was not possible to include entire chromosomes in the analyses. Instead, each chromosome was analyzed using overlapping windows of five to 15 SNPs at a time, considering all possible orders within a given window of markers. Recombination rate estimates were also obtained using CRIMAP.

In three instances we obtained linkage support for physical order, but the estimated recombination rates indicated that unlinked markers were present in that SNP window. In the first instance, a SNP (mCV24244050) on Chromosome 5 appeared completely unlinked across the majority of families. The likely explanation is that the SNP is placed at the end of the genetic map (near the centromere), where recombination events are difficult to establish. In the other two cases, SNPs in the middle of chromosomes 8 (rs13479764) and 12 (rs13481579) appeared unlinked in the map. The number of informative meioses for these two SNPs was low, and an examination of the distribution of crossovers indicated that they occurred in just two families. On Chromosome 8, the inflated recombination rates were observed only in the maternal meioses, whereas on Chromosome 12, the inflation was observed in the male meioses. On Chromosome 12, the

**Table 2.** Description of the SNP Genetic Map in the RIs

Chromosome	SNPs	Map Positions <sup>a</sup>	Genetic Length (cM)	Map Resolution <sup>b</sup>	Physical Length <sup>c</sup> (Mb)	cM/Mb
1	1,022	282	100.4	0.36	195.0	0.51
2	888	252	107.0	0.42	181.5	0.59
3	784	234	82.8	0.35	158.8	0.52
4	705	216	91.1	0.42	154.7	0.59
5	710	241	97.3	0.40	149.3	0.65
6	631	198	89.4	0.45	150.2	0.60
7	606	184	84.1	0.46	139.6	0.60
8	576	167	75.8	0.45	127.8	0.59
9	539	178	82.1	0.46	124.1	0.66
10	447	154	81.8	0.53	129.3	0.63
11	699	204	90.8	0.45	121.6	0.75
12	503	170	72.3	0.43	115.7	0.62
13	550	152	67.4	0.44	114.5	0.59
14	525	141	65.0	0.46	118.8	0.55
15	476	171	70.1	0.41	103.4	0.68
16	437	135	67.6	0.50	97.1	0.70
17	505	139	64.0	0.46	92.5	0.69
18	410	145	71.9	0.50	89.2	0.81
19	285	116	60.7	0.52	60.4	1.01
X	311	92	83.2	0.90	163.1	0.51
Total	11,609	3,571	1,604.8	0.45	2,586.6	0.62

<sup>a</sup>The number of unique positions of the map. Unique positions were considered as positions separated by more than 0.01 cM.

<sup>b</sup>The genetic length divided by the number of unique map positions.

<sup>c</sup>Physical length is the distance between the first and last SNP.

DOI: 10.1371/journal.pbio.0040395.t002

recombination rate was also inflated (but to a lesser extent) at a location eight markers upstream of the unlinked SNP. These inconsistencies may indicate the presence of chromosomal rearrangements. We excluded all these markers from the genetic maps.

Altogether we discarded 1,034 markers that mapped to multiple locations or did not map to any location on their assigned chromosome (see Materials and Methods). An additional 12 markers were discarded due to highly inflated recombination rates, which indicated inconsistencies in the map order. The remaining 10,202 markers had an average minor allele frequency of 0.26 (range: 0.003–0.5) and the

average physical distance between consecutive markers is 248 kb (Table S1).

The total length of the sex-averaged genetic map is 1,630 cM. The average genetic inter-marker distance is 0.16 cM, the largest inter-marker interval is 6.15 cM (between 53.6 and 59.8 cM) on Chromosome 5, and the average genetic distance between unique map positions (map resolution) is 0.37 cM for the sex-averaged map (Table 1). There are 59 regions with inter-marker interval above 2 cM. Sex-specific maps were also obtained in the HS data and, as observed in other species, showed pronounced variation in the rate of recombination. The lengths of the female and male genetic maps are 1,817 cM and 1,386 cM, respectively (Tables S2 and S3).

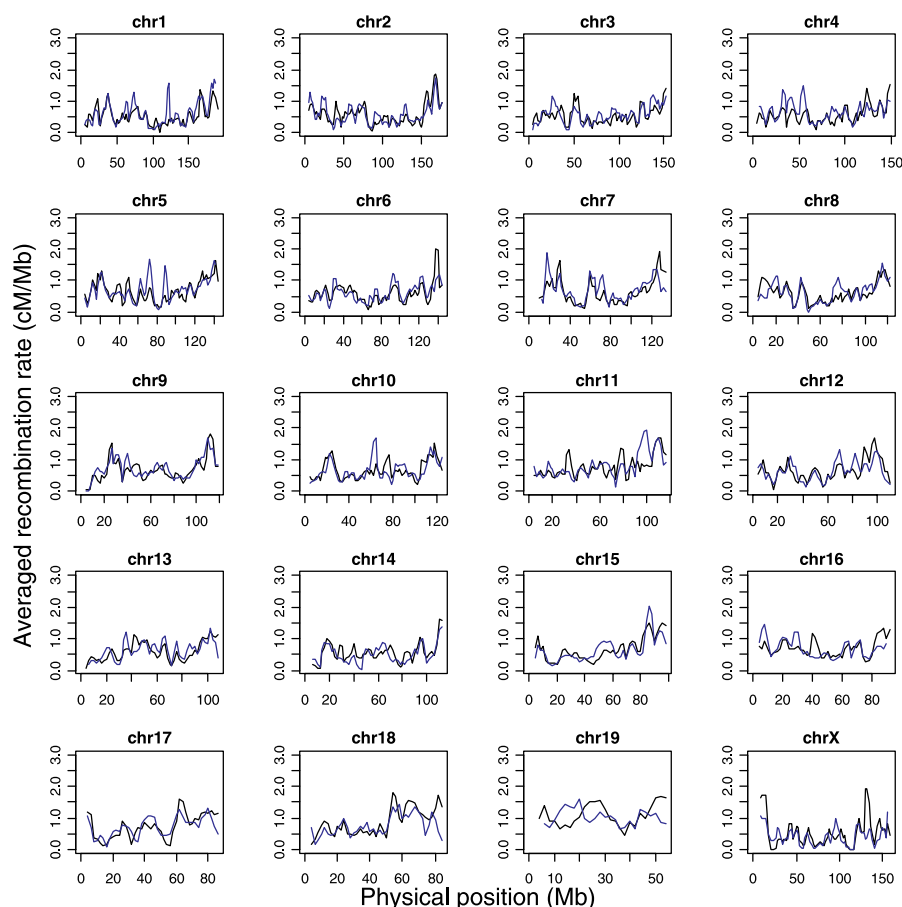
A consensus genetic map for recombinant inbreds was constructed using 11,609 informative SNPs and 267 strains from the eight sets of RIs. R/QTL was used to estimate the genetic distances between SNPs that were polymorphic in at least one of the RI sets [21]. On average, 52.6% of the 267 RI strains had informative genotypes per SNP (mean = 140 genotypes, SD = 60). Of the 273 RIs, six were not used for the map construction. AXB19, AXB20, AXB13, and BXA8 were excluded because their genotypes indicated that they could not be considered as independent strains [22,23]. AXB1 and AXB2 were excluded because of an extremely large number of flanking double recombination events (seven times more than other strains in the set). Twenty-three SNPs (0.2%) with three or more flanking double recombination events were excluded from the map construction. The genotypes of the rest of the SNPs are highly accurate: 124 genotypes (0.008% of genotypes) had an error log of odds (LOD) score of more than 3, which is consistent with the genotyping accuracy estimates of 99.99%. We also tested for segregation distortion from the expected 50% inheritance of each allele across all SNPs and

**Table 3.** Average Number of Recombinations for Different RI Sets

RI set	No. of Strains	Number of Polymorphic SNPs	Number of Recombinations	Recombinations per Strain
AKXD	24	5,042	1,027	42.8
AXB/BXA	34	7,696	1,848	54.4
Old BXD	36	7,367	1,733	48.1
New BXD	53	7,367	4,366	82.4
BXH	16	7,458	775	48.4
CXB	13	6,698	694	53.4
LXS	77	4,803	3,598	46.7
SWXJ	14	2,634	560	40.0
Total	267	11,609	14,601	54.7

Old BXD is the first developed set of strains [23,24], and new BXD is the set developed from advanced intercross populations [25].

DOI: 10.1371/journal.pbio.0040395.t003



**Figure 1.** Average Recombination Rates for All Chromosomes

Rates for RIs are shown as a black line and for HS as a blue line. The ratio between the genetic distance and physical distance was calculated using a sliding window of 5 Mb and a shift of 2 Mb between windows centers, assuming a constant rate of recombination between two adjacent markers. DOI: 10.1371/journal.pbio.0040395.g001

RI strains. The combined  $p$ -value (Fisher method for combining  $p$ -values) across all strains, corrected for the number of SNPs, did not reach significance for any of the SNPs.

In the BXD set, 52 SNPs showed variation in genotypes that corresponded to the three different phases of development of the BXD RIs [24–26] (Table S4). Forty-seven SNPs are not polymorphic in the 26 BXD strains established from a single cross of a C57BL/6J female to a DBA/2J male, but are polymorphic in similar BXD strains established more than 20 y later. Five SNPs are not polymorphic in the first 36 BXD strains, but are polymorphic in the newest set of 53 BXD lines (BXD43–100). The 52 SNPs are not polymorphic in other inbred strains examined, including highly related strains such as DBA/1J and C57BL/10J. These data indicate that the 52 SNPs are recent mutations that have arisen since the creation of the BXD lines. Twenty-four of the mutations are in the DBA/2J strain and 28 are in C57BL/6J. For the map purposes, we treated the BXD as three different sets for the analysis of the 52 SNPs. The genotypes for the non-polymorphic subset were treated as missing values.

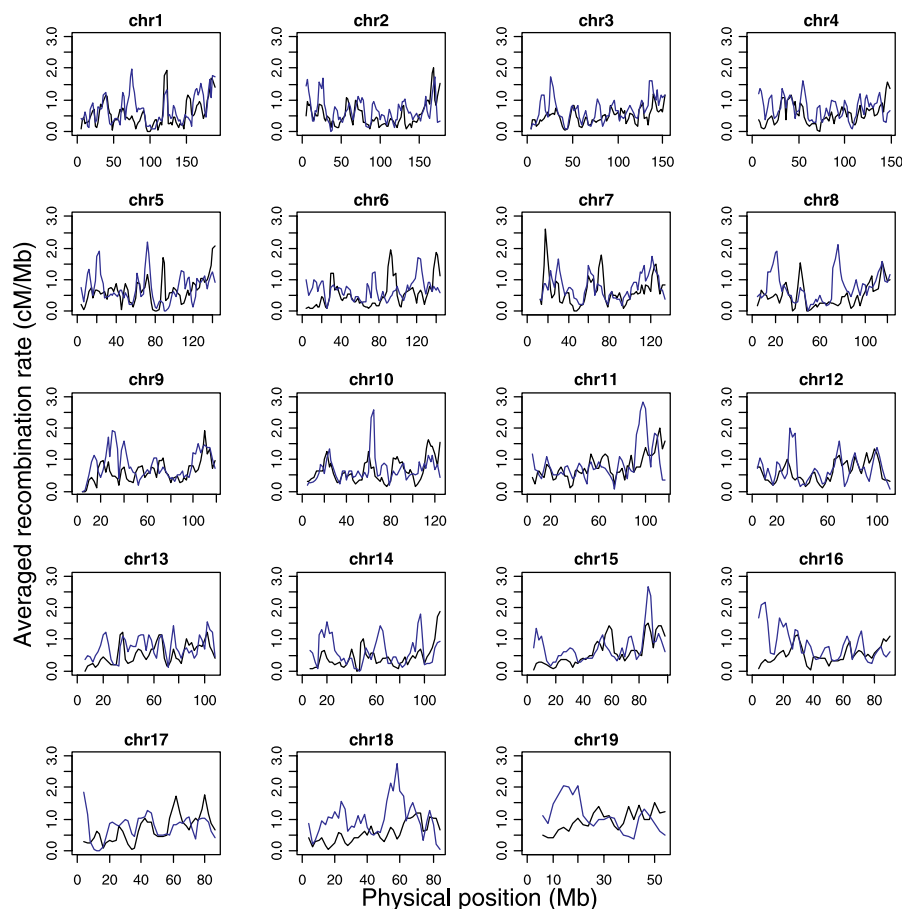
We converted RI recombination frequencies to the distances expected for a single-generation meiotic map. The total genetic length of the RI map from proximal to distal SNP on each chromosome is 1,605 cM (Table S5). The genetic lengths of each chromosome are presented in Table 2, and

the number of recombinants in each set is shown in Table 3. One BXD subset was developed using advanced intercross populations [26] and therefore contains approximately 1.7 times more recombinants per line. The average number of recombinations in each of the other sets is correlated with the number of polymorphic SNPs in the set (Spearman correlation coefficient  $r = 0.84$ ,  $p = 0.02$ ). The average distance between recombination events (map length divided by the number of recombinations) in the RI map is 0.11 cM (29.3 cM per strain) or 177.1 kb (47.3 Mb per strain). The average genetic distance between SNPs in the RI map is 0.14 cM or 216.4 kb, and the average resolution is 0.45 cM (Table 2).

### Recombination Rates

We compared genetic and physical distances and calculated recombination rates, expressed as centimorgans per megabase across the genome. The genome-wide average recombination rate is 0.63 cM/Mb for the HS sex-averaged map and 0.62 cM/Mb for the RI map. The rates varied between chromosomes and range from 0.43–0.88 and 0.51–1.01 cM/Mb per chromosome for the HS and RIs, respectively.

Variation in recombination rates between chromosomes is partly accounted for by chromosome size (HS:  $r = -0.60$ , proportion of variance explained  $R^2 = 0.35$ ,  $p = 0.0056$ ; RIs:  $r = -0.85$ ,  $R^2 = 0.73$ ,  $p = 1.6 \times 10^{-6}$ ). There is a negative



**Figure 2.** Sex-Specific Recombination Rates in the HS

Average recombination rate for all chromosomes calculated similar to the HS sex-averaged map, for males (black line) and females (blue line).

DOI: 10.1371/journal.pbio.0040395.g002

correlation between chromosome size and chromosome average recombination rate, with the recombination rate of the smallest chromosome (Chromosome 19) being approximately 1.5-fold higher than that of the longest chromosome (Chromosome 1) in the HS and 2-fold higher in the RIs. The correlation between chromosome size and recombination rate is higher in female HS ( $r = -0.61$ ,  $p = 0.0041$ ) than in male HS maps ( $r = -0.47$ ,  $p = 0.044$ ). The X chromosome does not account for this difference between females and males because it exhibits a lower rate of recombination relative to the rate expected for its size. The correlation between chromosome size and recombination rate in females excluding chromosome X is nearly the same ( $r = -0.61$ ,  $p = 0.0054$ ).

We calculated recombination rates along chromosomes using non-overlapping windows of 1, 3, and 5 Mb. These window sizes were chosen as a compromise between the limits of the marker resolution (about 200 kb) and the need to sample sufficient recombination events within a window. The correlation between recombination rates calculated for each window for the HS and RI maps is 0.50, 0.60, and 0.66, respectively ( $p < 1 \times 10^{-60}$  for all windows) (Figure 1). The range of recombination rates for 1-, 3-, and 5-Mb windows varies from 0–6, 0–3, and 0–2 cM/Mb, respectively, with higher rates of recombination near the telomeres.

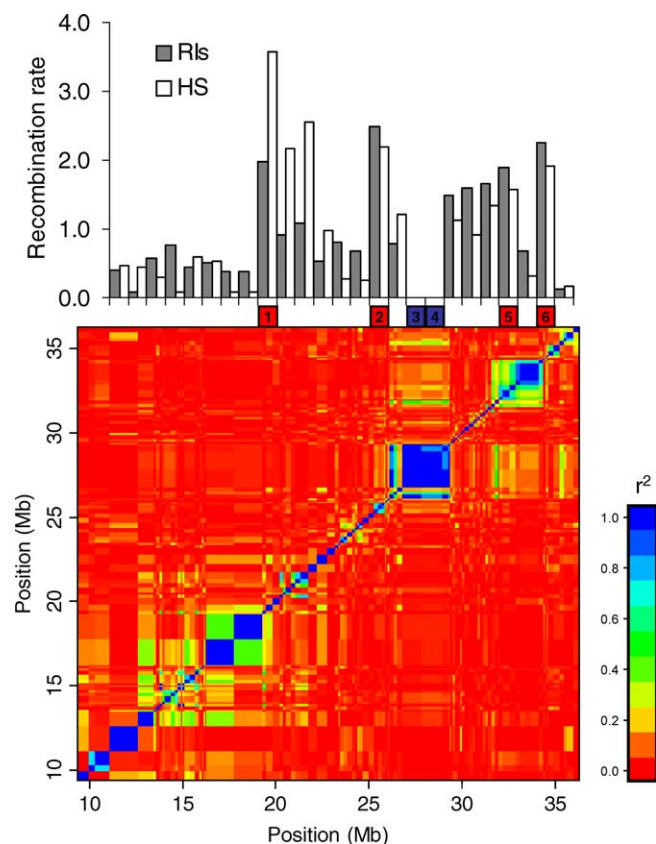
We calculated recombination rates in males and females separately using the sex-specific HS map for all autosomes.

On average, the autosomal rate of recombination is higher in females (0.72 cM/Mb) than in males (0.57 cM/Mb). The recombination rate in females is notably higher near the centromere, whereas in males it is higher near the telomere (a highly significant difference:  $p = 2.3 \times 10^{-7}$ ), similar to the pattern observed in humans. We also examined the correlation between male and female recombination rates and identified many regions showing high recombination rate in one sex, but not the other (Figure 2). For example, there are four sex-specific peaks on Chromosome 6 between 60 Mb and the end of the chromosome. The correlation between sex-specific recombination rates calculated using 1-, 3-, and 5-Mb non-overlapping windows for males and females are 0.28, 0.32, and 0.33, respectively ( $p < 1 \times 10^{-6}$ ).

### Correlation of Recombination with Sequence Parameters

To investigate the association between recombination and sequence features, we analyzed megabase-sized regions with extreme rates of recombination in both HS and RIs. A 1-Mb non-overlapping window was used to identify the top and bottom 10% of intervals with the highest and lowest recombination rates (termed here recombination “jungles” and “deserts,” respectively). Out of the 494 extreme regions, 154 are recombination jungles and deserts in both HS and RIs (84 recombination jungles and 70 recombination deserts). Thirteen regions are discordant, being in the high 10% based





**Figure 3.** Relation between LD Levels in Inbred Strains and Recombination Rates in HS and RIs on Chromosome 7

The top figure is the recombination rate (cM/Mb) for the HS and RIs in 1-Mb non-overlapping windows across a 25-Mb region in Chromosome 7. Below, the six boxes represent the location of four recombination jungles (in red) and two deserts (in blue) of 1 Mb each. The bottom figure is the LD structure of the same region based on the SNP genotypes of 55 inbred strains. Each dot represents the pairwise LD level between two different SNPs. Regions of high LD measured by  $r^2$  are in blue, whereas regions of linkage equilibrium are in red. As can be seen in this example, deserts tend to overlap with large blocks of LD, and jungles are typically located on the border of the blocks. The overlap proportion of the four recombination jungles with haplotype blocks is 0.00, 0.58, 0.55, and 0.35 for regions 1, 2, 5, and 6. The two deserts (3 and 4) overlap completely with haplotype blocks.

DOI: 10.1371/journal.pbio.0040395.g003

on one map (in ten of 13 cases in the HS map) and in the low 10% based on the other map.

To test whether jungles and deserts are conserved, we defined the haplotype block structure of the mouse genome using 55 inbred strains that were genotyped for the same SNP panel. As in human populations, haplotype blocks were defined to be regions where there has been little historical recombination [27]. We investigated the relations between recombination rates and LD structure by comparing the position of jungles and deserts with the boundaries of haplotype blocks (Figure 3). We identified 682 haplotype blocks with an average size of 0.90 Mb (range: 0.0086–10.23 Mb), spanning a total of 611.56 Mb (23% of the genome). To compare block boundaries with the recombination jungles and deserts regions, we aligned the blocks with the jungles and deserts, and for each 1-Mb region, examined the proportion within haplotype blocks. An overlap of more than 80% between the haplotype blocks boundaries and a

jungle or a desert was considered a region that predominantly lies within haplotype blocks. Fifty-nine percent of the total span of all recombination deserts overlap with haplotype blocks compared to only 12% for the jungles ( $p = 2.4 \times 10^{-13}$ ). Only one jungle predominantly lies within a haplotype block in contrast to 33 (47.1%) of the deserts. Nevertheless, there are 19 (25.7%) deserts that do not correspond to haplotype blocks.

To search for factors that predict recombination jungles and deserts, we first examined the association between recombination rate and global sequence features. We found differences in the abundance of many sequence variables, including simple repeats, interspersed repetitive sequences, and transposable elements (Table 4), but gene and SNP density were not significantly enriched in jungles or deserts. We found a higher abundance of simple repeats in recombination jungles, including (CA)<sub>n</sub> repeats (relative risk [RR] = frequency in jungles/frequency in deserts = 1.8), (TG)<sub>n</sub> repeats (RR = 1.9), and low-complexity repeats (RR = 1.4). In contrast, (TA)<sub>n</sub> repeats (RR = 0.79) and some long interspersed nuclear elements (LINEs) (RR = 0.6) are under-represented in recombination jungles, as seen also in the study of human recombination hotspots [16]. An individual LINE repeat (termed L1<sub>MM</sub>), enriched in recombination deserts, is the most significant feature associated with recombination (RR = 0.3,  $p = 2.8 \times 10^{-35}$ ).

We examined the seven nucleotide sequence motif (CCTCCCT) previously shown to be enriched in human recombination hotspots [16]. We compared the number of occurrences of this sequence in recombination jungles and deserts. The CCTCCCT motif is significantly enriched in recombination jungles (RR = 1.9,  $p = 6.8 \times 10^{-29}$ ), and the number of motifs is a very strong predictor of recombination jungles versus deserts (Figure 4A): it explains 63% of the deviance in a logistic regression model. In addition to CCTCCCT, we also tested a nine nucleotide oligomer (CCCCACCCC) that was also identified in human recombination hotspots. The CCCCCACCCC motif is also enriched in recombination jungles (RR = 2.2,  $p = 2.2 \times 10^{-20}$ ), but it does not significantly add to the deviance explained in a logistic regression model. Looking at recombination rate across the whole genome, the correlation between the CCTCCCT motif and the average recombination rate in non-overlapping windows of 1 Mb is 0.35 ( $p = 8.8 \times 10^{-71}$ ).

We tested for the effect of sex on the association of the CCTCCCT motif with recombination. We used the HS sex-specific recombination rates calculated in larger windows (2 Mb) to make the estimates comparable to those obtained from sex-averaged estimates. The CCTCCCT motif was found to be associated with recombination rates in both males and females, with correlation of 0.35 and 0.26, respectively (males,  $p = 1.5 \times 10^{-35}$ ; females,  $p = 1.1 \times 10^{-18}$ ). We did not find associations between other sequence elements and recombination rates that differed between the sexes.

Since our findings suggest that the role of the CCTCCCT motif in recombination is conserved across different mammalian species, we asked whether the position of the motif is conserved between species. We identified the orthologous regions corresponding to the 154 regions of mouse jungles and deserts in the genomes of the rat, dog, chimpanzee, and human. The density of the CCTCCCT motif (counts per Mb) is highly correlated between all species across the different

**Table 4.** Repeat Elements and Sequence Variables Associated with Recombination Jungles and Deserts

Sequence Feature	Sequence Motif <sup>a</sup>	Recombination Deserts		Recombination Jungles		RR <sup>b</sup>	p-Value
		Mean	SD	Mean	SD		
<b>DNA repeat elements</b>	CCTCCCT	188.9	43.9	350.7	89.2	1.9	$6.8 \times 10^{-29}$
	CCCCACCCC	31.9	13.1	82.6	37.7	2.6	$2.2 \times 10^{-20}$
	All	18.3	15.2	50.5	18.2	2.8	$1.2 \times 10^{-23}$
	URR1A	3.2	2.9	8.2	4.6	2.5	$1.4 \times 10^{-13}$
	URR1B	3.2	2.6	8.0	4.1	2.5	$2.1 \times 10^{-15}$
<b>Simple repeat</b>	All	324.7	81.6	458.9	89.1	1.4	$1.1 \times 10^{-17}$
	(TA) <sub>n</sub>	31.8	14.3	20.7	9.0	0.7	$1.3 \times 10^{-7}$
	(CA) <sub>n</sub>	41.5	14.3	76.6	23.6	1.8	$1.4 \times 10^{-21}$
	(TG) <sub>n</sub>	41.0	13.6	76.5	23.7	1.9	$4.4 \times 10^{-22}$
<b>LINE</b>	All	348.6	93.9	221.9	73.5	0.6	$8.4 \times 10^{-16}$
	L1_MM	84.4	22.0	28.0	18.9	0.3	$2.8 \times 10^{-35}$
<b>SINE</b>	All	268.7	167.7	612.1	339.1	2.3	$2.9 \times 10^{-13}$
	B4A	21.3	17.0	56.3	33.8	2.6	$1.3 \times 10^{-13}$
<b>Low complexity</b>	All	167.9	52.0	137.4	26.9	0.8	$2.4 \times 10^{-5}$
	AT_RICH	126.6	46.8	73.2	27.7	0.6	$1.9 \times 10^{-13}$
	CT-RICH	11.2	5.9	19.5	7.5	1.7	$2.0 \times 10^{-12}$
	GA-RICH	11.9	9.0	20.3	6.2	1.7	$8.9 \times 10^{-10}$
<b>LTR</b>	All	216.7	63.7	198.7	52.6	0.9	$6.0 \times 10^{-2}$
	MTD	12.7	6.4	23.3	11.9	1.8	$8.7 \times 10^{-11}$
<b>%GC</b>		60.0	4.1	55.3	3.3	0.9	$2.1 \times 10^{-12}$
<b>CpG</b>		3.9	2.9	8.5	8.2	2.2	$6.4 \times 10^{-6}$

<sup>a</sup>Repeat classes are in bold followed by individual elements within the class that are significantly associated with recombination deserts or jungles.

<sup>b</sup>Ratio of frequency in jungles to frequency in deserts.

LTE, long terminal repeat; SINE, short interspersed nuclear element.

DOI: 10.1371/journal.pbio.0040395.t004

regions (correlation between mouse and rat: 0.84, dog: 0.67, chimpanzee: 0.67, and human: 0.74, maximum observed  $p = 5.5 \times 10^{-20}$ ). To see how these cross-species correlations in the frequency of the CCTCCCT motif correspond to recombination rates, we calculated the recombination rate (cM/Mb) for the mouse orthologous regions in humans using the Rutgers high-resolution genetic map. Recombination rates in humans are on average higher in regions orthologous to mouse jungles compared to regions orthologous to deserts ( $p = 1.6 \times 10^{-5}$ ), and correlate significantly with the mouse rates: the correlations are 0.35 ( $p = 2.0 \times 10^{-5}$ ) and 0.42 ( $p = 2.2 \times 10^{-7}$ ) with the values in RIs and HS, respectively. Figure 4B plots the frequency of recombination rates for the human regions that are orthologous to mouse jungles and deserts.

## Discussion

We have constructed one of the most dense and accurate genetic maps for any organism. Using both outbred animals and recombinant inbred lines, we genetically placed more than 10,000 markers onto the mouse genome, yielding an average genetic inter-marker distance of 0.15 cM. The average resolution is 0.37–0.45 cM/Mb for the sex-averaged maps. Even small numbers of genotype errors will inflate genetic map distances, and this factor has often made it difficult to draw conclusions about the factors that influence local recombination rates. Our high level of genotyping accuracy (~99.99%), the use of large pedigrees, and the construction of separate maps based on two types of populations has result in a highly reliable linkage map and highly accurate estimates of recombination rates.

Our map was constructed using HS mice that were also employed to find QTLs [11]. We collected 101 phenotypes for

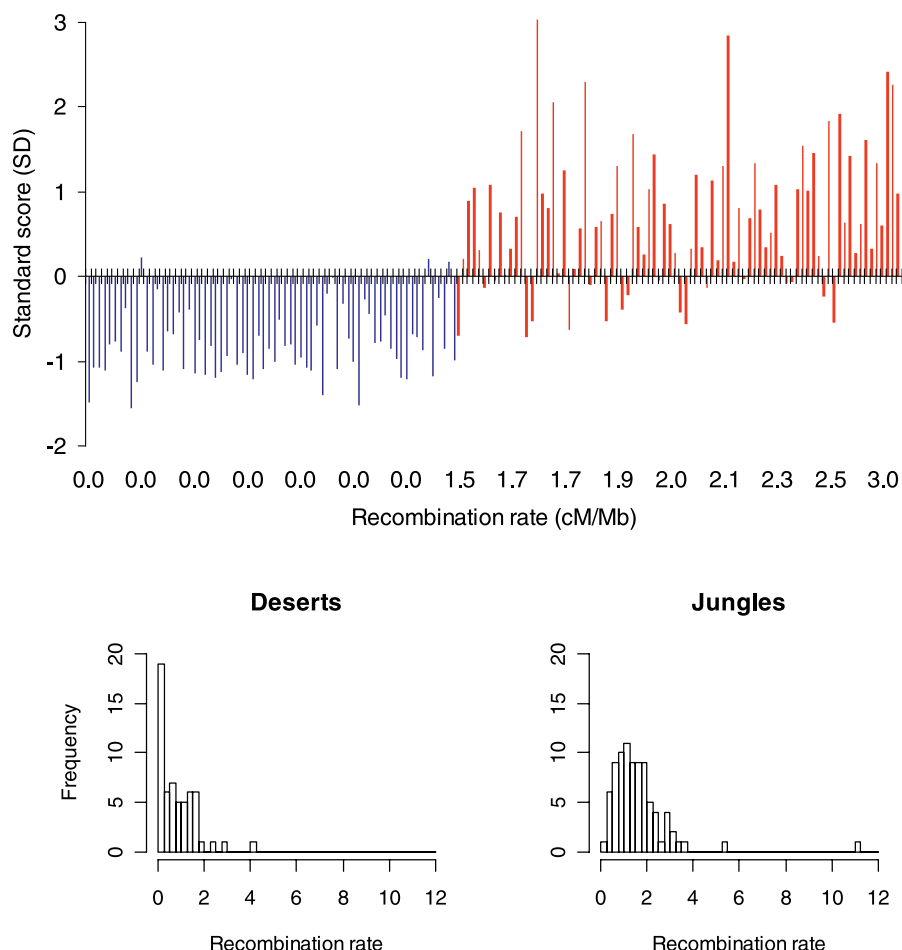
1,904 HS mice (389 parents of these animals were also genotyped but not phenotyped, accounting for the 2,293 HS animals used in map construction). For the QTL analysis, 10,202 markers unambiguously positioned on the HS were used, supplemented by a further 1,910 informative markers whose locations were derived by interpolation from the physical map then available (NCBI build 34). Thus a sex-averaged map of 12,112 markers was used for QTL identification [11]. We have not yet explored the effect on QTL identification of using sex-specific maps.

We found that chromosome-specific recombination rates do not greatly differ between outbred (HS) and inbred populations (RIs): there are very few regions with discrepant recombination rates. Petkov et al. [28] suggest that inbreeding results in strong selection for specific allele combinations in mouse inbred lines: selection against recombination occurs when a combination of alleles at linked loci confers a survival advantage during the inbreeding process. Our data do not support the hypothesis of lower recombination events in inbreds or of strong selection forces that favor specific alleles. The inheritance of alleles in the RIs is not significantly different from what is expected.

We do see, however, some differences related to chromosome size. Recombination in the RIs appears to be more constrained by chromosome size, which explains 73% of the variance in the chromosome average recombination rate, whereas in the HS, it explains only 35% of the variance. Higher recombination rates are observed in the HS relative to the RIs in the large chromosomes, and lower rates in the small chromosome (with the exception of the X chromosome). We cannot at this point explain this observation.

We found remarkable similarity in the features that contribute to variation in recombination rates in mouse





**Figure 4.** Recombination Jungles and Deserts

Enrichment of CCTCCCT motif in recombination jungles (upper panel). The standard deviation (SD) from the mean number of CCTCCCT motifs is plotted against recombination rate. Red bars are 1-Mb regions included in the top 10% recombination rate in both HS and RIs (recombination jungles), and blue bars are regions in the bottom 10% (recombination deserts).

Histograms of human recombination rates for the orthologous regions of the mouse recombination jungles and deserts (lower panel).

DOI: 10.1371/journal.pbio.0040395.g004

and human chromosomes. We confirmed that chromosome, position on the chromosome, sex, and sequence composition are common important factors. In both species, differences in recombination rates are correlated with chromosome size (or arm size) and proportional distance from the centromere. It is possible that, as has been suggested in humans, variation in recombination rates between chromosomes occurs because each chromosome arm is constrained to have a single chiasma, and the effect of fixing the lower limit of recombination will affect smaller chromosome arms more than larger [29,30]. Unlike human chromosomes, mouse chromosomes are all acrocentric (they lack a fully-formed P arm). Humans therefore have approximately twice as many chromosomal arms as mice, which would explain the difference in average recombination rates: our estimate of 0.6 cM/Mb is about half that found in humans. However, data from rats do not support this notion [31].

In both humans and mice, the average recombination rate in females is higher than males. Recombination rates are higher near the centromeres for females and towards the telomeres for males [32,33]. The correlation between female and male recombination rates across the whole genome is

relatively weak, with many sex-specific peaks and troughs in recombination rates. This suggests that a larger proportion of the variance in recombination can be explained by sex-specific recombination events. A direct measure of recombination using immunohistological methods has shown that sex, and not genotype, influences recombination [34].

The main difference between the LD landscapes of the human and mouse genome is the presence of large haplotype blocks of up to several megabases in size in mouse inbred strains. We have shown that the position of these haplotype blocks tends to coincide with recombination deserts. In contrast, recombination jungles coincide with borders of these blocks or regions without evident block structure. An important implication of this observation is that genetic variants (such as those underlying quantitative trait loci) that lie within a recombination desert will be difficult to identify with the SNP maps we have generated.

We exploited the LD landscape of the mouse genome to concentrate on regions of very high-recombination rates (jungles) and very low rates (deserts). We used the selected jungles and deserts, as well as the recombination rates in the whole genome, to study sequence features that were shown to

influence global and local recombination rates in humans. A major limitation of this analysis is that the location of recombination hotspots within jungles is unknown. Although deserts are essentially homogenous regions without recombination, jungles are presumably heterogeneous regions with an uneven distribution of recombination events. Therefore, in this study we focused on factors that are known to influence recombination in humans.

In humans, sequence motifs are an important determinant of variation in recombination rate. Analyses of the sequence correlates of recombination based on low resolution genetic maps have been unable to identify motifs that independently predict recombination rate. Most of the correlations have low coefficients and the sequence variables are correlated with each other. High resolution mapping of recombination hotspots in humans found CCTCCCT oligomers to be the strongest signal of recombination hotspots [16]. The motif is also the strongest predictor of recombination jungles in the mouse, explaining 63% of the deviance. In humans, this motif is found within the long terminal repeats of two retrovirus-like transposons, THE1A and B. The mouse genome does not contain this motif, providing additional support for its independent effect.

Our data support a two-stage model of recombination in which recombination rates are constrained over large scales, but are rapidly evolving on a small scale. The conservation of recombination deserts in inbred lines, shown by haplotype blocks and the presence of some of those deserts also in humans, are also consistent with this model. Comparisons between species using low-resolution sex-averaged maps have previously detected only a slight positive correlation between recombination rates [31]. The fact that the frequency of the CCTCCCT motif is well correlated across species in the regions that we studied, whereas the location of hotspots is not conserved, suggests that an interaction of several elements, including the CCTCCCT motif, influences recombination. The high degree of similarity between the factors that are correlated with mouse and human recombination rates, together with the correlation in recombination rates itself, implies the existence of a common mechanism that influences recombination rates in mammalian genomes. The genetic maps we have made are available from [http://gscan.well.ox.ac.uk/#genetic\\_map](http://gscan.well.ox.ac.uk/#genetic_map) and as Tables S1–S5.

## Materials and Methods

**SNP selection and genotyping.** We selected SNPs across the genome that distinguish between the eight HS founders (A/J, AKR/J, BALB/c, DBA/2J, C57BL/6J, LP/J, I, and RIIS/J). We used datasets to select SNPs that are validated and polymorphic in at least some of the HS founders [6,19,35]. SNPs were chosen on the basis of their position in the genome, regardless of any known, or putative, function. SNPs were merged together and remapped onto the mouse genome (NCBI build 34). SNPs closer than 50 kb with identical strain distribution patterns (SDP) were discarded. All gaps without SNPs over 500 kb were filled using SNPs from Celera, Affymetrix SNPs, and from a comparative study of CZECH/Ei. All genotyping was carried out at Illumina (San Diego, California, United States) using the BeadArray genotyping platform. PEDSTATS [36] and PedCheck [37] were used to identify non-Mendelian transmissions in the HS data. The Wellcome–CTC mouse strain SNP genotypes are available at <http://www.well.ox.ac.uk/mouse/INBREDS>.

**Genetic map construction: HS.** The mapping algorithm used the physical sequence positions to determine the initial order of the markers on the map, followed by estimation of linkage support for the physical order, and maximum-likelihood estimates of the recombination fractions. The majority of markers were present in NCBI build 34 of the mouse genome, and where these were

unavailable (125 SNPs), build 33 locations were used to interpolate marker order on build 34. We used CRIMAP [20] to construct linkage maps applying the Kosambi map function [38].

Linkage support for physical order was obtained in two stages. The first step was to confirm localization of each marker to the chromosome assigned by its reported physical location on NCBI build 34 of the mouse genome. In the set of 11,247 informative markers, we required each marker to exhibit linkage to at least one other marker on that chromosome, with odds greater than 1000:1, using the *tuopoint* option in CRIMAP. Altogether, 11,199 markers were linked to at least one marker on the chromosome with a LOD > 3, and 48 markers were discarded at this stage. The second test of map order consisted of the removal and remapping of each marker onto the map. A marker was kept in its physical NCBI build 34 location only if the likelihood of the physical location was within 1 LOD unit of the highest likelihood obtained.

Because of the complexity of some of the pedigrees and the number of markers, it was not possible to include entire chromosomes in the analyses. Therefore, each chromosome was analyzed using overlapping windows of five to 15 SNPs at a time, by considering all possible orders within a given window of markers using the *all* function in CRIMAP. If a marker did not map to within 1 LOD unit of its physical location, we searched for the highest likelihood of placement along the remaining windows for that chromosome. At this stage, 10,213 markers mapped to their physical location using this method. The remaining 986 markers mapped to multiple locations along the designated chromosome.

Once linkage support for marker order was established, recombination rate estimates were obtained for the 10,213 markers in CRIMAP. Recombination rate estimates were calculated using the *fixed* function of CRIMAP in a series of overlapping windows of five to 15 SNPs using the maximum number of estimated data points (over all windows) to calculate the recombination fraction between each pair of markers. Following recombination rate estimation, two markers in the map had inflated recombination rates and appeared completely unlinked in all windows analyzed. Because the density of markers indicated that these were highly unlikely events, we excluded these markers from the maps, and excluded one window on Chromosome 12 for which two such events occurred within a window of nine consecutive markers. As a result, 10,202 markers were placed on the map.

In the final set of 10,202 markers, we estimated double recombinant events, which may indicate erroneous genotypes or potential chromosomal rearrangements. We estimate the rate at which two or more recombinant events occurred within non-overlapping windows of five markers using the *chrompic* function in CRIMAP. The resulting estimate over the genome was  $2.3 \times 10^{-4}$ , taking into account the average number of informative meioses per chromosome. However, we did not exclude any genotypes at this stage on the premise that the double recombinant events may represent chromosomal rearrangements, in particular, deletions.

**Genetic map construction: RIs.** The RI sets include: AXB (22 individuals, 7,696 polymorphic SNPs), BXA (18 individuals, 7,696 polymorphic SNPs), CXB (13 individuals, 6,698 polymorphic SNPs), BXD (89 individuals, 7,367 polymorphic SNPs), BXH (16 individuals, 7,458 polymorphic SNPs), AKXD (24 individuals, 5,042 polymorphic SNPs), LXS (77 individuals, 4,803 polymorphic SNPs), and SWXJ (14 individuals, 2,634 polymorphic SNPs).

Within each set of RI animals, we deleted any non-polymorphic SNPs from each set, as well as SNPs whose flanking sequences were not mapped to a single location in the NCBI build 34 mouse genome assembly. The genotypes were recorded as 1 or 2, based on the alleles of the founders, and heterozygotes were treated as missing values. The SNPs were ordered based on build 34. The number of double recombinations of adjacent SNPs was calculated for each SNP and each strain. If more than two double recombinants were found for a SNP interval, it was taken to indicate potential genotype errors or errors in the order of markers. Markers with more than two double recombinations were excluded. For the genetic distance estimation, the genotypes of the eight RIs sets from 267 strains were combined into one dataset. Alleles of SNPs that are not polymorphic in an RI set were treated as missing data. The error LOD score [39] was calculated for each genotype to identify possible genotyping errors using the *R/QTL* [21] function *calc.errorlod*. *R/QTL* was used to estimate the genetic distances between 11,609 SNPs that were polymorphic in at least one of the RI sets.

**Haplotype blocks.** Inbred strains were genotyped for the same panel of SNPs with the Illumina platform as mentioned above. Haplotype blocks were defined using a 95% confidence bound on *D'* (the confidence interval algorithm) using Haploview 3.0. The following 55 inbred strains were selected from 72 inbred strains

after removing 17 strains that are in high correlation ( $>0.8$ ) with other strains: 129S6/SvEv, 129X1/SvImJ, A/J, AKR/J, ALR/LtJ, ALS/LtJ, BALB/cByJ, BPH/2J, BPL/1J, BPN/3J, BTBR T+ tf/J, BUB/BnJ, C3HeB/FeJ, C57BL/10J, C57L/J, C58/J, CALB/RkJ, CBA/J, CE/J, DBA/1J, DDY/JclSeyFrk, EL/SuzSeyFrk, FVB/NJ, I/LnJ, ILS, IS/CamRkJ, ISS, KK/HlJ, LEWES/EiJ, LG/J, LP/J, MA/MyJ, NON/LtJ, NOR/LtJ, NZB/BINJ, NZO, NZW/LacJ, PERA/EiJ, PERC/EiJ, PL/J, RBA/DnJ, RBF/DnJ, RF/J, RIIS/J, SB/LeJ, SEA/GnJ, SENCARA/PtJ, SF/CamEiJ, SJL/J, SM/J, ST/bJ, SWR/J, WMP/PasDnJ, WSB/Ei, and YBR/Ei.

**Sequence features.** We obtained gene and repeat features from the Ensembl database (data derived from mouse assembly NCBI build 34) using standard Perl modules [40]. GC content and the presence of particular sequence motifs were obtained by directly scanning DNA sequences of interest using bespoke Perl scripts. SNP density was calculated based on the SNPs from the Broad Institute (<http://www.broad.mit.edu/snp/mouse/>) together with SNPs from Perlegen (<http://mouse.perlegen.com/mouse/download.html>). We used the Student *t*-test to compare the abundance of features in recombination jungles and deserts. Logistic regression modeling was performed using the glm function in the R statistical analysis package version 2.1.1, with the type of region (jungle vs. desert) as the response, and the frequency of the sequence motif in the region as explanatory variable. Goodness of fit was tested using the model's deviance (twice the log likelihood ratio). Orthologous regions were defined between mouse (Build34 mm6, University of California Santa Cruz Genome Browser [<http://genome.ucsc.edu/>] nomenclature) and other assembled genomes (human, hg17; rat, rn3; dog, canFam1; and chimpanzee, panTro1) using the Genome Browser "Net" and "Chain" alignments as described by Taylor et al. [41].

## Supporting Information

**Table S1.** A High-Resolution Genetic Map of the Mouse Genome

The map gives the sex-averaged genetic position of 9,904 SNP markers from 2,293 heterogeneous stock mice. The columns are the marker name (marker), the genome build of the markers' physical position (genome\_\_build), the chromosome, base pair (bp) position, and genetic position (cM).

Found at DOI: 10.1371/journal.pbio.0040395.st001 (830 KB XLS).

**Table S2.** Male-Specific Genetic Map of the Mouse Genome Based on HS Pedigrees

The columns are the marker name (marker), the genome build of the markers' physical position (genome\_\_build), the chromosome, base pair (bp) position, and genetic position (cM).

Found at DOI: 10.1371/journal.pbio.0040395.st002 (830 KB XLS).

**Table S3.** Female-Specific Genetic Map of the Mouse Genome Based on HS Pedigrees

The columns are the marker name (marker), the genome build of the

markers' physical position (genome\_\_build), the chromosome, base pair (bp) position, and genetic position (cM).

Found at DOI: 10.1371/journal.pbio.0040395.st003 (976 KB XLS).

**Table S4.** SNPs That Distinguish Subsets of BXD Strains

The columns are the marker name (marker), the chromosome, base pair (bp) position, the genotype of the BXD parental strains C57BL/6J (B) and DBA/2J (D), the genotypes of two highly related strains C57BL/10J and DBA/1J, the genotypes of the BXD lines sorted on the basis of the three different phases of development. B indicates an allele from C57BL/6J, D indicates alleles from DBA/2J, and H is heterozygote.

Found at DOI: 10.1371/journal.pbio.0040395.st004 (88 KB XLS).

**Table S5.** A High-Resolution Genetic Map Obtained from Analysis of Eight Sets of RI Lines

The columns are the marker name (marker), the genome build of the markers' physical position (genome\_\_build), the chromosome, base pair (bp) position, and genetic position (cM)

Found at DOI: 10.1371/journal.pbio.0040395.st005 (1.1 MB XLS).

## Acknowledgments

We thank members of the Complex Trait Consortium for providing DNA samples and funding to support the genotyping of RI strains: Beth Bennett, Gary A. Churchill, Jing Gu, Kent Hunter, Lu Lu, Richard S. Nowakowski, and Petko Petkov. Many thanks to Tim Wiltshire, Mathew Pletcher, Eric Schadt, Petko Petkov, Mark Daly and Andrew Kirby, Weikuan Gu, and Christophe Benoist for providing SNP information. We also thank Tara Matise and Chunsheng He for helpful discussions about HS genetic maps. We thank Karl Broman for help with the RI analysis.

**Author contributions.** RWW, RM, and JF conceived and designed the experiments. SS, JTB RRC, MST, and RM analyzed the data. RWW and RM contributed reagents/materials/analysis tools. SS, JTB, and JF wrote the paper. SS and JTB performed data cleaning and statistical analysis. RRC and MST performed the sequence analysis. RW coordinated sample collection. RM and JF formulated and directed the project.

**Funding.** This work was supported by the Wellcome Trust. SS is supported by a European Molecular Biology Organization (EMBO) fellowship. Genotyping of RI strains was supported by multiple grants from the National Institutes of Health (National Institute of Neurological Disorders and Stroke, National Institute on Alcohol Abuse and Alcoholism, National Institute of General Medical Sciences, National Institute on Drug Abuse, National Cancer Institute, and National Institute of Mental Health).

**Competing interests.** The authors have declared that no competing interests exist.

## References

- Dietrich WF, Miller J, Steen R, Merchant MA, Damron-Boles D, et al. (1996) A comprehensive genetic map of the mouse genome. *Nature* 380: 149–152.
- European Backcross Collaborative Group (1994) Towards high resolution maps of the mouse and human genomes—a facility for ordering markers to 0.1 cM resolution. *Hum Mol Genet* 3: 621–627.
- Rhodes M, Straw R, Fernando S, Evans A, Lacey T, et al. (1998) A high-resolution microsatellite map of the mouse genome. *Genome Res* 8: 531–542.
- Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, et al. (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature* 420: 520–562.
- Lindblad-Toh K, Winchester E, Daly MJ, Wang DG, Hirschhorn JN, et al. (2000) Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. *Nat Genet* 24: 381–386.
- Wiltshire T, Pletcher MT, Batalov S, Barnes SW, Tarantino LM, et al. (2003) Genome-wide single-nucleotide polymorphism analysis defines haplotype patterns in mouse. *Proc Natl Acad Sci U S A* 100: 3380–3385.
- Pletcher MT, McClurg P, Batalov S, Su AI, Barnes SW, et al. (2004) Use of a dense single nucleotide polymorphism map for in silico mapping in the mouse. *PLoS Biol* 2: e393. DOI: 10.1371/journal.pbio.0020393
- Moran JL, Bolton AD, Tran PV, Brown A, Dwyer ND, et al. (2006) Utilization of a whole genome SNP panel for efficient genetic mapping in the mouse. *Genome Res* 16: 436–440.
- Crawford DC, Bhargale T, Li N, Hellethel G, Rieder MJ, et al. (2004) Evidence for substantial fine-scale variation in recombination rates across the human genome. *Nat Genet* 36: 700–706.
- McVean GA, Myers SR, Hunt S, Deloukas P, Bentley DR, et al. (2004) The fine-scale structure of recombination rate variation in the human genome. *Science* 304: 581–584.
- Valdar W, Solberg LC, Gauguier D, Burnett S, Klennerman P, et al. (2006) Genome-wide genetic association of complex traits in heterogeneous stock mice. *Nat Genet* 38: 879–887.
- Talbot CJ, Nicod A, Cherny SS, Fulker DW, Collins AC, et al. (1999) High-resolution mapping of quantitative trait loci in outbred mice. *Nat Genet* 21: 305–308.
- Darvasi A, Soller M (1995) Advanced intercross lines, an experimental population for fine genetic mapping. *Genetics* 141: 1199–1207.
- Daw EW, Thompson EA, Wijsman EM (2000) Bias in multipoint linkage analysis arising from map misspecification. *Genet Epidemiol* 19: 366–380.
- Fingerlin TE, Abecasis GR, Boehnke M (2006) Using sex-averaged genetic maps in multipoint linkage analysis when identity-by-descent status is incompletely known. *Genet Epidemiol* 30: 384–396.
- Myers S, Bottolo L, Freeman C, McVean G, Donnelly P (2005) A fine-scale map of recombination rates and hotspots across the human genome. *Science* 310: 321–324.
- Winckler W, Myers SR, Richter DJ, Onofrio RC, McDonald GJ, et al. (2005) Comparison of fine-scale recombination rates in humans and chimpanzees. *Science* 308: 107–111.
- Demarest K, Koyner J, McCaughan J Jr., Cipp L, Hitzemann R (2001) Further characterization and high-resolution mapping of quantitative trait loci for ethanol-induced locomotor activity. *Behav Genet* 31: 79–91.
- Cervino AC, Li G, Edwards S, Zhu J, Laurie C, et al. (2005) Integrating QTL and high-density SNP analyses in mice to identify *Insig2* as a susceptibility gene for plasma cholesterol levels. *Genomics* 86: 505–517.

20. Lander ES, Green P (1987) Construction of multilocus genetic linkage maps in humans. *Proc Natl Acad Sci U S A* 84: 2363–2367.
21. Broman KW, Wu H, Sen S, Churchill GA (2003) R/qtl: QTL mapping in experimental crosses. *Bioinformatics* 19: 889–890.
22. Sampson SB, Higgins DC, Elliot RW, Taylor BA, Lueders KK, et al. (1998) An edited linkage map for the AXB and BXA recombinant inbred mouse strains. *Mamm Genome* 9: 688–694.
23. Williams RW, Gu J, Qi S, Lu L (2001) The genetic structure of recombinant inbred mice: High-resolution consensus maps for complex trait analysis. *Genome Biol* 2: research0046. 1–0046. 18.
24. Taylor BA (1978) Recombinant inbred strains: Use in gene mapping. In: Morse HC III, editor. *Origins of inbred mice: Proceedings of a workshop*, Bethesda, Maryland, February 14–16, 1978. New York: Academic Press. pp. 423–438.
25. Taylor BA, Wnek C, Kotlus BS, Roemer N, MacTaggart T, et al. (1999) Genotyping new BXD recombinant inbred mouse strains and comparison of BXD and consensus maps. *Mamm Genome* 10: 335–348.
26. Peirce JL, Lu L, Gu J, Silver LM, Williams RW (2004) A new set of BXD recombinant inbred lines from advanced intercross populations in mice. *BMC Genet* 5: 7.
27. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, et al. (2002) The structure of haplotype blocks in the human genome. *Science* 296: 2225–2229.
28. Petkov PM, Graber JH, Churchill GA, DiPetrillo K, King BL, et al. (2005) Evidence of a large-scale functional organization of mammalian chromosomes. *PLoS Genet* 1: e33. DOI: 10.1371/journal.pgen.0010033
29. Laurie DA, Hulten MA (1985) Further studies on chiasma distribution and interference in the human male. *Ann Hum Genet* 49 (Pt 3): 203–214.
30. Pardo-Manuel de Villena F, Sapienza C (2001) Recombination is proportional to the number of chromosome arms in mammals. *Mamm Genome* 12: 318–322.
31. Jensen-Seaman MI, Furey TS, Payseur BA, Lu Y, Roskin KM, et al. (2004) Comparative recombination rates in the rat, mouse, and human genomes. *Genome Res* 14: 528–538.
32. Broman KW, Murray JC, Sheffield VC, White RL, Weber JL (1998) Comprehensive human genetic maps: Individual and sex-specific variation in recombination. *Am J Hum Genet* 63: 861–869.
33. Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, et al. (2002) A high-resolution recombination map of the human genome. *Nat Genet* 31: 241–247.
34. Lynn A, Schrupp S, Cherry J, Hassold T, Hunt P (2005) Sex, not genotype, determines recombination levels in mice. *Am J Hum Genet* 77: 670–675.
35. Petkov PM, Cassell MA, Sargent EE, Donnelly CJ, Robinson P, et al. (2004) Development of a SNP genotyping panel for genetic monitoring of the laboratory mouse. *Genomics* 83: 902–911.
36. Wigginton JE, Abecasis GR (2005) PEDSTATS: Descriptive statistics, graphics and quality assessment for gene mapping data. *Bioinformatics* 21: 3445–3447.
37. O'Connell JR, Weeks DE (1998) PedCheck: A program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* 63: 259–266.
38. Kosambi DD (1944) The estimation of map distances from recombination values. *Ann Eugen* 12: 172–175.
39. Lincoln SE, Lander ES (1992) Systematic detection of errors in genetic linkage data. *Genomics* 14: 604–610.
40. Potter SC, Clarke L, Curwen V, Keenan S, Mongin E, et al. (2004) The Ensembl analysis pipeline. *Genome Res* 14: 934–941.
41. Taylor MS, Kai C, Kawai J, Carninci P, Hayashizaki Y, et al. (2006) Heterotachy in mammalian promoter evolution. *PLoS Genet* 2: e30. DOI: 10.1371/journal.pgen.0020030